

Per Computer zu den Anfängen der Evolution

„Stammbäume des Lebens“: Am Heidelberger HITS wird Grundlagenforschung zur Berechnung der Artenbildung betrieben

Von Arndt Krödel

Seine Spezialität sind Stammbäume, allerdings nicht im herkömmlichen Sinn familiärer Ahnenforschung: Der Informatiker Dr. Alexandros Stamatakis, Forschungsgruppenleiter am Heidelberger Institut für Theoretische Studien (HITS), entwickelt Programme, mit denen sich extrem große Stammbäume verschiedenster Organismen am Computer berechnen lassen – Rekonstruktionen evolutionärer Zusammenhänge durch Hochleistungsrechner. Stamatakis (Foto: Kreuzer) arbeitete zusammen mit Forschern aus den USA an einem Projekt, das den bisher größten Stammbaum der Pflanzen berechnete: Der „BigPlantTree“ umfasst exakt 55 473 Arten von Blütenpflanzen, so genannte Angiospermen, und deckt damit etwa 90 Prozent aller Pflanzenarten der Welt ab – ein Meilenstein in der evolutionären Bioinformatik.



Pflanzen sind für den Wissenschaftler aber nur ein Anwendungsfall seiner rechnergestützten Erforschung der stammesgeschichtlichen Entwicklung der Lebewesen, in der Fachsprache Phylogenese (von gr. phylon = Stamm, genesis = Entstehung) genannt. Auch Mikroorganismen wie Bakterien oder Viren oder auch Wirbeltiere können Forschungsgegenstand sein. Selbst die Entstehung der Sprachen im Verlauf der Entwicklung der Menschheit wird durch die phylogenetische Methode untersucht. Eine spezielle Software für eine Spezies gibt es nicht: „Es ist immer das gleiche Berechnungsmodell, das möglichst global anwendbar sein soll“, erläutert Stamatakis. „Das entspricht eigentlich dem Ziel der Informa-

tion, immer möglichst allgemeingültige, abstrakte Lösungen für eine große Klasse von Problemen zu erarbeiten“.

Um die evolutionären Zusammenhänge von Organismen bis zu den Wurzeln zu verfolgen, verwenden die Wissenschaftler in erster Linie die DNA-Sequenzanalyse: Per Computer wird die Abfolge und Position der Spezies im Stammbaum von charakteristischen Abschnitten in einem Strang der DNA, der Trägerin der Erbinformation des Lebewesens, bestimmt. Die so gewonnenen Erkenntnisse über die verwandtschaftlichen Zusammenhänge der verschiedenen Arten stellen die Forscher in einem „phylogenetischen Stammbaum“ dar wie im BigPlantTree-Projekt. So etwas geht natürlich nicht am Laptop: „Da braucht man schon eine große Zahl leistungsfähiger Prozessoren, die alle gemeinsam an der Berechnung dieses einen großen Stammbaums arbeiten“, sagt Stamatakis. Im Instituts-Rechenzentrum, das er neben seiner Forschungsgruppe leitet, verfügt der neue Parallelrechner über rund 2 000 Prozessoren.

Wissenschaftlern stehen die am HITS entwickelten Programme frei zur Verfügung – jeder Interessierte kann sich die Software herunterladen und seine Datensätze dann selbstständig analysieren. Dr. Stamatakis versteht sich hier als Brückenbauer zwischen Evolutionsbiologie und dem Hochleistungsrechnen. Den Schwerpunkt seiner wissenschaftlichen Arbeit sieht er in der Weiterentwicklung und Verfeinerung der Software, gerade vor dem Hintergrund der Menge an genetischen Daten, die durch neue Se-



Der Purpursonnenhut ist ein leuchtendes Beispiel aus dem Stammbaum der Blütenpflanzen, den Dr. Alexandros Stamatakis und sein Team auf ihrem Rechner bereithalten. Foto: HITS

quenzierungstechniken stark ansteigt. „Dieser Datenwelle rennen wir eigentlich immer hinterher“, umreißt der Informatiker das Dilemma, dass die Anzahl der molekularbiologischen Informationen schneller wächst als die Leistungsfähigkeit moderner Prozessoren.

Im Moment arbeitet er mit seinen Kollegen an einem noch größeren Stammbaum der Pflanzen mit ungefähr 100 000 Arten. Nur ein Parallelrechner wie der im HITS-Rechenzentrum vorhandene kann die anfallenden Datenmengen von 70 Gigabyte (70 Milliarden Byte) bewältigen. Die Forschungsgruppe tüftelt an einer Lösung, die es demnächst ermöglichen wird, den Speicherverbrauch des Riesenbaums auf nur fünf Gigabyte zu reduzieren. „Damit

könnte auch ein kleines Labor mit einem normalen Server einen solchen Stammbaum berechnen“, stellt der Wissenschaftler fest.

Nächste große Herausforderung ist die Einbindung in ein Projekt, bei dem die Stammbäume von Vögeln berechnet werden sollen, wofür das jeweilige Genom – die Gesamtheit der Erbinformation – von etwa 100 Arten sequenziert wird. Spannend daran könnte zum Beispiel sein, anhand des Baumes festzustellen, wie sich der Gesang einzelner Vögel entwickelt hat oder zu welchem Zeitpunkt eine neue Flugcharakteristik entwickelt wurde. Unterschiedliche Ausmaße von Artenbildung könnten Hinweise auf einen Zusammenhang mit einem Ereignis in der Erdgeschichte sein, etwa einer Klimaänderung oder einem Meteoriteneinschlag.